What is claimed is:

1. A computerized storage and retrieval system of biological information comprising:

a means for data entry;

a means for displaying the data;

a programmable central processing unit for performing automated analysis; and

a data storage means containing protein pathways and annotated information on the pathways stored in a relational database, wherein the pathways annotated and organized in a curated clustering arrangement and wherein the annotated information is accessed through the relational database.

2. The computer system of claim 1, wherein the information pertaining to the pathways is stored in a plurality of tables further comprising proteins, their sequences and attributes; protein interactions; protein-protein associations; protein pathways; mRNA, microarray, and protein expression data; genes, their sequences and attributes; and descriptions of cells, tissues, organs, pathology reports, patient histories, and treatments.

3. The computer system of claim 1, wherein the central processing unit is programmed to retrieve, input, edit, annotate, search, calculate similarities, align, and predict homologous or orthologous protein pathways.

4. The computer system of claim 1, wherein the central processing unit is programmed to perform protein sequence analysis, protein interactions analysis, protein-protein association analysis, protein pathway analysis, gene expression analysis, pathway annotation analysis, pathway edit analysis, pathway expression analysis, tissue expression analysis, subtractive hybridization analysis, electronic northern analysis, or commonality analysis.

5. The computer system of claim 1, wherein the data is entered using the standard for pathway representation.

6. The computer system of claim 1, wherein a means for displaying the data is used to show two related pathways as a diagram containing nodes which represent proteins or non-protein molecules; modes that represent protein interactions or protein-protein associations; scores calculated from sequence, motif or structural homologies that interrelate nodes; and coefficients of similarity that interrelate modes of the pathway.

7. The computer system of claim 1, wherein the central processing unit is programmed to compare two protein pathways by a node-only, a mode-only, or a node-and-mode comparison and wherein the node-only comparison is selected from protein only, non-protein only, and protein and non-protein nodes.

8. The computer system of claim 1, wherein the central processing unit is programmed to run an algorithm for dynamic programming comprising:

a) initializing an array, in which a two dimensional array $M=M_{ij}$ with J rows and variant length for each row, the length for i-th row is n_i is set up and $M_{Ji}=0$, where $1<=i<=n_J$,

b) backfilling the array via backward recursion with the formula

$$M_{ik} = \max_{\substack{j>i \\ i\le l\le n_j}} \left\{ w\left(a_{ik}, a_{jl}\right) + M_{jl}\theta\left(w\left(a_{ik}, a_{jl}\right)\right)\right\} \text{ for } 1 \le k \le n_i,\ 1 \le i \le J$$

where θ(.) is the step function defined as θ (v)={0, if v<=0; 1, if v>0} and w(.,.) is the scoring function between the two nodes, defined as

$$w(a_{ik}, a_{jl}) = \begin{cases} 0, \text{ if i=j, } a_{ik}=a_{jl},\ a_{ik}=-D, \text{ or } a_{jl}=-D \\ \theta\left(c_{ik,jl}-t_c\right)\cdot\left\{\alpha\left(1-\left|s_{ik}-s_{jl}\right|\right)+(1-\alpha)c_{ik,jl}\right\} \text{ otherwise.} \end{cases}$$

, and D>0

c) using traceback to identify putative pathways PPW_j, 1<=j<=max n_i with the top *n* best scores.

9. A method for performing pathway editing comprising programming the central processing unit of claim 1 to identify interactions among proteins; weigh the interactions; and calculate coefficients of similarity for the interactions, thereby producing an OS score and editing the protein pathway.

10. A method of using genes which encode known proteins to annotate modes of a protein pathway comprising:

a) using the computer system of claim 1 to select genes which encode known proteins,

b) employing the genes to produce a protein-protein association matrix containing coefficients of similarity, and

c) annotating the modes of the pathway using the coefficients of similarity from the matrix.

11. A method for protein pathways analysis using a node-and-mode comparison comprising:

a) submitting a query pathway and protein sequences; and

b) allowing the computer system of claim 1 to

i) compare nodes using the dynamic programming algorithm wherein a sequence identity score or p-value summarizes similarity and wherein a weighting factor between 0 and 1 is assigned to corresponding nodes,

ii) compare modes by generating a SCIM matrix, thereby assigning a coefficient of similarity to corresponding modes,

iii) align pathways globally or locally, wherein insertion or deletion of nodes or modes incurs a penalty,

iv) sum all similarity scores, and

v) display at least one high-scoring segment of the aligned pathways.

12. A method for performing protein pathways analysis comprising:

a) submitting a query pathway and protein sequences; and

b) allowing the computer system of claim 1 to

i) organize and analyze the query pathway and protein sequences,

ii) compare protein sequence identity of the query with all protein sequences in the protein pathways database using standard methods of protein comparison,

iii) use a SCIM matrix to derive and compare coefficients of similarity for each interaction of the query and all interactions for proteins in the protein pathways database,

iv) calculate an OS-score based on sequence identity and coefficients of similarity,

remove all pathways not meeting user-specified threshold for OS-score, and

vi) retrieve aligned pathways meeting the threshold.

13. A method for searching a protein pathways database for protein interactions comprising:

a) submitting a query pathway;

b) allowing the central processing unit of claim 1 to perform protein interactions analysis between the query pathway and all protein pathways in the protein pathways database wherein coefficient of similarity is produced to interrelate each mode of the query pathway and a mode of the most closely related protein pathway; and

c) retrieving at least one protein pathway alignment.

14. A method of using a query pathway to search a protein pathways database to predict homologous pathways comprising:

a) submitting a query pathway and protein sequences;

b) allowing the central processing unit of claim 1 to compare the query pathway and protein sequences with all protein pathways and proteins in the protein pathways database, and

c) retrieving a plurality of pathway alignments wherein the homologous pathways are aligned by OS-score.

15. A method of using a known protein pathway and a protein database to predict orthologous pathways comprising:

a) submitting a query pathway and known protein sequences,

b) allowing the central processing unit of claim 1 to compare known sequences to all protein sequences stored in the database,

c) retrieving orthologous proteins with the highest identity to the known proteins,

d) inheriting protein interactions from the query pathway, and

e) aligning the query pathway and the orthologous proteins, thereby predicting orthologous pathways.

16. A method of using a known protein pathway to predict the nodes and modes of a novel pathway comprising:

a) submitting a query pathway and known protein sequences;

b) applying standard methods of comparison to determine similarity between the known protein sequences and protein sequences in the protein databases, thereby predicting candidate nodes;

c) utilizing coefficients of similarity from protein interactions or protein-protein association data, thereby predicting candidate modes; and

d) retrieving novel pathways with an OP-score obtained using an optimization algorithm.

17. The method of claim 16, wherein coefficients of similarity are based on mRNA/cDNA counting, microarray expression, protein expression, known protein-protein associations, a promoter similarity matrix, or more than one of these methods.

18. The method of claim 16, further comprising using a constrained clustering method wherein the clustering method is average linkage, single linkage, complete linkage, K-means, or self-organizing maps; the constraint is that no more than one protein in each cluster is derived from a single column of aligned proteins; and the accuracy of the prediction is determined by an OP-score.

19. A method for predicting novel pathways comprising:

a) generating candidate proteins from one species for each node based on a protein search;

b) employing a means for optimization to find likely linear linkages between candidate proteins aligned to the query pathway with possible gaps in the alignment, and

c) reporting all pathways with optimal and sub-optimal predictions that satisfy user-specified alignment and interaction parameters wherein the accuracy of the prediction is provided by OP-score.

20. The method of claim 19, wherein the means for optimization is based on linear next-neighbor criteria, global minimization criteria, dynamic programming, or iterative searches using at least two of the means.

21. A method for determining the function of a protein or a gene that encodes the protein comprising:

a) placing the protein encoded by the gene in a candidate pathway involving at least two proteins, and

b) using the data storage means of claim 1 wherein the interactions with proteins and non-protein molecules, cellular location, and expression are used to determine the function of the protein or gene.

22. A method for predicting novel pathways comprising:

a) submitting a query pathway and protein sequence

b) using the computer system of claim 1 to process the query pathway and protein sequences using orthologous pathway prediction wherein the data is derived from protein similarities and interactions, or homologous pathway prediction wherein the data is derived from protein similarities and interactions, from protein-protein associations, and

c) applying a dynamic programming algorithm or a constrained clustering algorithm, thereby predicting the novel pathways.